Comparative interrater reliability of Asian Stroke Disability Scale, modified Rankin Scale and Barthel Index in patients with brain infarction

__Kavian Ghandehari__[(1)], Kosar Ghandehari[(2)], Ghazaleh Saffarian-Toosi[(3)],
Shahram Masoudinezhad[(3)], Siamak Yazdani[(3)], Ali Nooraddin[(3)],
Saeed Ebrahimzadeh[(4)], Fahimeh Ahmadi[(3)], Fatemeh Abrishamchi[(3)]

Abstract

BACKGROUND: This study tried to develop an Asian Stroke Disability Scale (ASDS) and compared its interrater reliability with modified Rankin Scale (mRS) and Barthel Index (BI).

METHODS: Three items including self-care, mobility, and daily activities were selected as variables for development of the ASDS. The variables were provisionally graded on a 2- to 4-point scale based on the importance of each item. Each of the variables was categorized into 3 categories. Afterward, 125 rater-patient assessments for each scale (mRS, BI, and ASDS) were performed on 25 stroke patients by 5 raters. For categorization of functional impairment as minor or major, the scores of mRS, BI and ASDS were categorized as ≤ 2 and > 2, < 90 and ≥ 90, and < 3 and ≥ 3, respectively.

125 rater-patient assessments for each of the mRS, BI, and ASDS were performed on 25 stroke patients by five raters.

RESULTS: The quantitative variability of BI, mRS, and ASDS scores was not significant ($P = 0.379$; $P = 0.780$; and $P = 0.835$, respectively). Interrater variability of mRS, BI, and ASDS scores based on qualitative categorization was not significant ($P = 1.000$; $P = 0.978$; and $P = 0.901$, respectively). Paired interrater variability of mRS, BI, and ASDS scores based on qualitative categorization was not significant ($P > 0.05$).

CONCLUSION: The ASDS is easy to use, requires less than 1 minute to complete and is as valid as mRS and BI in assessment of functional impairment of patients with stroke.

Introduction

Many studies have assessed of quality of life and outcomes after stroke. In the US, Duncan et al. found that eight key areas (strength, hand function, activities of daily living, mobility, communication, memory, emotion, and social participation) emerged as key areas from patients' perspective.[1] Similarly, Williams et al. reported that patients identified 12 key domains (mobility, energy, upper extremity function, work/productivity, mood, self-care, social roles, family roles, vision, language, thinking, and personality).[2] Feeding, grooming, dressing, bathing and toileting are considered as basic activities of daily living.[3] On the other hand, shopping, using transportation, telephoning, preparing meals, handling finances, work skills, and leisure time activities are among instrumental activities of daily living.[3]

Since a reproducible and valid method for quantification of the neurological deficit that occurs after stroke is essential for monitoring patients, many stroke scales have been proposed.[4] Stroke scales are useful tools for estimating the severity of stroke at onset and for assessment of prognostic information in hospital. In general, a stroke scale consists of several variables to observe signs and symptoms. Each variable is categorized for scoring.[4] Modified Rankin Scale (mRS) and Barthel Index (BI) are the most commonly used tools to measure disability and handicap after stroke. While acceptability and average completion time need to be considered in designing any outcome measure, they are particularly important when stroke outcomes are assessed since the patients may have cognitive problems and feelings of tiredness after stroke.[5]

1- Professor, Cerebrovascular Research Center, Neuroscience Research Center, Department of Neurology, School of Medicine Mashhad University of Medical Sciences, Mashhad, Iran
2- Research Fellow, Cerebrovascular Research Center, Mashhad University of Medical Sciences, Mashhad, Iran
3- Resident, Department of Neurology, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran
4- Lecturer, Department of Biostatistics, Mashhad University of Medical Sciences, Mashhad, Iran
Correspondence To: Kavian Ghandehari, Email: kavianghandehari@yahoo.com

Stroke scales can be classified as parametric or clinimetric scales which measure physical deficit and functional scales which evaluate functional recovery after stroke.[4,6] The popular National Institute of Health Stroke Scale (NIHSS) is categorized as a clinimetric scale, while the commonly used mRS and BI are categorized as functional impairment scales. Japan Stroke Scale and Kurashiki Prehospital Stroke Scale are Asian clinimetric stroke scales.[4,7] On the other hand, Chinese Stroke Scale is a comprehensive Asian functional impairment scale.[8] We designed the Asian Stroke Disability Scale (ASDS) as a simplified functional impairment and handicap scale and compared its interrater reliability with mRS and BI.

Materials and Methods

Consecutive stroke patients admitted in Ghaem Hospital (Mashhad, Iran) were enrolled in a cross-sectional study in September 2011. Cognitive impairment, language problems, stroke due to subarachnoid hemorrhage, hepatic failure, renal failure, age less than 18 years old, and decreased level of consciousness were considered as exclusion criteria.[9] Patients who were unable to take care of their own affairs prior to stroke were also excluded.[9] Stroke was defined as a sudden focal neurologic deficit of presumed vascular origin lasting for at least 24 hours with or without corresponding lesion on brain imaging.[10] The procedure for developing the ASDS can be summarized as follows: 1) selecting variables, 2) categorizing the variables, 3) evaluating the categories for their distribution and sensitivity, 4) modifying and reevaluating the categorization, 5) repeating steps 1-4 until the appropriate categorizations are obtained.[4]

Based on the contribution of each item to the prognosis of stroke and a review of currently available stroke scales,[4] three items including self-care, mobility, and daily activities were selected as variables for development of the ASDS. According to the importance of each item, the variables were provisionally graded on 2- to 4-point scales. Each variable was categorized into three categories which were expressed in a concrete way, avoiding abstractive expression. Therefore, the same grade could be obtained regardless of the level of training of the rater. Patients' total scores ranged from zero to eight and were calculated by summing the scores of all variables (Table 1).

Five residents of neurology assessed the scores of patients on mRS, BI, and ASDS seven days after the stroke.[11] In order to eliminate any potential training effects, we only included the last 25 patients with stroke which resulted in 125 rater-patient assessment for each tool (mRS, BI, and ASDS).[12] Each assessor performed score detection based on the mRS, BI, and ASDS while no other assessor was present in the room. Moreover, all assessors were blinded to ratings of others.[13]

For categorization of functional impairment as minor or major, the scores of mRS, BI, and ASDS were categorized as ≤ 2 and > 2, < 90 and ≥ 90, and < 3 and ≥ 3, respectively.[14] Reliability is a measure of consistency in multi-item scales. Quantifying the reproducibility of scoring between graders gives interobserver variability for our analysis. Variability was described using parametric one-way analysis of variance (ANOVA) for BI. Non-parametric Mann-Whitney U test and Kruskal-Wallis H test were used for quantitative evaluation of mRS and ASDS based on the results of Kolmogorov-Smirnov test. For qualitative categorization of mRS, BI, and ASDS, variability was analyzed using chi-square and Fisher's exact tests. The research was approved by the local ethics committee and informed consents were obtained from the patients.

Results

The mean age of patients was 65.5 years (range: 45-85 years). Females constituted 44% of the participants. The final diagnosis was ischemic stroke in 68% and hemorrhagic stroke in 22%. In quantitative evaluation of BI by all five raters, median BI score was 57.520 (standard deviation: 27.567; standard error: 2.466). Interrater variability of BI scores was not significant (df = 4; F = 1.061; 95% confidence interval = 52.639-62.400; P = 0.379). In quantitative evaluation of mRS of all five raters, the mean mRS score was 3.416 ± 1.277. Interrater variability of mRS scores was not significant (X2 = 1.758; df = 4; P = 0.780). The mean ASDS score in quantitative evaluation of the five raters was 4.760 ± 2.357. Interrater variability of ASDS scores was not significant (X2 = 1.454; df = 4; P = 0.835).

Table 2 shows paired interrater variability of mRS and ASDS scores in our study group based on the Mann-Whitney U test. The paired interrater variability of mRS, BI, and ASDS scores based on qualitative categorization was not significant (Table 3). Interrater variability of scores of mRS (X2 = 0.553; df = 4; P = 1.000), BI (X2 = 0.869; df = 4; P = 0.978), and ASDS (X2 = 1.434; df = 4; P = 0.901) based on the qualitative categorization was not statistically significant.

K. Ghandehari, K. Ghandehari, Gh. Saffarian-Toosi, Sh. Masoudinezhad, S. Yazdani, A. Nooraddin, S. Ebrahimzadeh, F. Ahmadi, F. Abrishamchi

Table 1. The Asian Stroke Disability Scale (ASDS)

Mobility (chair to bed, walking, stairs)
0- No problems, independent at all items
1- Some problems, needs walker or help of another person
2- Sever problems, wheelchair, immobile , bedridden

Self-care (feeding, toileting, dressing, bathing, grooming)
0- No problems with self-care, independent at all items
1- Some problems, needs help
2- Unable or totally dependent

Daily activities (work and social, transport, family, leisure, sex, and recreational activities)
0- No problems with daily activities due to stroke
2- Some problems
4- Unable

Note: Scores are calculated based on the difference between before and after stroke.

Table 2. Paired interrater quantitative variability of modified Rankin Scale (mRS) and the Asian Stroke Disability Scale (ASDS) in 25 patients with stroke

Code of raters	mRS		ASDS	
	Z	P	Z	P
1-2	-0.062	0.951	-0.276	0.782
1-3	-0.886	0.376	-0.356	0.722
1-4	-0.322	0.748	-0.197	0.844
1-5	-1.104	0.270	-0.955	0.340
2-3	-0.657	0.511	-0.530	0.596
2-4	-0.267	0.789	-0.363	0.716
2-5	-0.903	0.367	-1.096	0.273
3-4	-0.615	0.539	-0.137	0.891
3-5	-0.083	0.933	-0.589	0.556
4-5	-0.804	0.421	-0.685	0.493

Table 3. Paired interrater variability of Barthel Index (BI), modified Rankin Scale (mRS) and the Asian Stroke Disability Scale (ASDS) scores based on qualitative categorization of 25 patients

Code of raters	mRS (≤ 2 and > 2)		BI (< 90 and ≥ 90)		ASDS (< 3 and ≥ 3)	
	X^2	P	X^2	P	X^2	P
1-2	0.222	1.000	0.596	0.702	0.595	0.702
1-3	0.000	1.000	0.166	1.000	0.166	1.000
1-4	0.000	1.000	0.166	1.000	0.166	1.000
1-5	0.222	1.000	0.000	1.000	0.122	0.463
2-3	0.222	1.000	0.136	1.000	0.136	1.000
2-4	0.222	1.000	0.136	1.000	0.136	1.000
2-5	0.000	1.000	0.595	0.702	0.117	1.000
3-4	0.000	1.000	0.000	1.000	0.000	1.000
3-5	0.222	1.000	0.166	1.000	0.500	0.725
4-5	0.222	1.000	0.166	1.000	0.500	0.725

Discussion

Evaluating the impact of new treatments requires the use of reliable and valid outcome measures.[15] In contrast to BI, distinction between grades of mRS are poorly defined.[16] However, the instruments and approaches developed to date have not consistently been shown to reduce interrater variability for mRS.[13,16] MRS and BI are widely used stroke disability scales which have been proven to be valid and reliable for defining outcome in patients with stroke.[14] Therefore, we compared the interrater variability of ASDS with mRS and BI which are assumed as gold standard. The quantitative and qualitative interrater variability of ASDS in were similar to those of mRS and BI. Quinn et al. found interrater reliability of mRS to be poor (k = 0.16). They also compared estimated scores between paired assessors and suggested poor agreement in 30% of mRS scores and significant variability (k = 0.38).[12] In evaluation of mRS, raters' scores concurred fully in 47 of 50 patients and in the remaining 3 patients, scores differed by 1 level.[13]

A review of literature about interrater reliability of mRS revealed moderate interrater reliability which improved with structured interviews.[17] Cincura et al. showed that using structured interviews for completing mRS improves interobserver concordance rates. They reported NIHSS, BI, and mRS to have good validity when translated and culturally adopted.[18]

Testing intraobserver variability over a short time period will be biased by observer recall of previous grading. On the other hand, delaying the second testing allows for potentially significant patient improvement or disease progression.[19] Therefore, we did not assess intrarater reliability in our validation study. The differences between disability scores on mRS, BI, and Scandinavian Stroke Scale are small and these scores have excellent agreement with each other. Modified NIHSS was also found to have substantial agreement with mRS and BI in a study in UK.[20] Another comparative study in UK was performed on 1400 patients. When mRS and BI scores were dichotomized at 95 and 1, respectively, NIHSS appeared more sensitive than BI or mRS.[21]

Diagnostic accuracy of BI in serial assessments of patients with ischemic stroke was evaluated in the Netherlands. BI scores early after stroke showed good discriminative properties for final outcome of BI at 6 months.[11] Another study in the Netherlands compared four stroke scales including Orgogozo Scale, NIHSS, the Canadian Neurological Scale, and the Scandinavian Stroke Scale with measures of disability and handicap and quality of life according to mRS and BI. The five stroke scales were highly related to one another but the correlation between stroke scales and functional scales was less than 0.70 and decreased from BI (47.5%) to mRS (36.5%).[22] Finally, the ASDS is easy to use, requires less than 1 minute to complete and is as valid as mRS and BI in assessment of functional impairment of patients with stroke.

Conflict of Interests

Authors have no conflict of interests.

References

1. Duncan PW, Wallace D, Lai SM, Johnson D, Embretson S, Laster LJ. The stroke impact scale version 2.0. Evaluation of reliability, validity, and sensitivity to change. Stroke 1999; 30(10): 2131-40.
2. Williams LS, Weinberger M, Harris LE, Clark DO, Biller J. Development of a stroke-specific quality of life scale. Stroke 1999; 30(7): 1362-9.
3. Kelly-Hayes M, Robertson JT, Broderick JP, Duncan PW, Hershey LA, Roth EJ, et al. The American Heart Association Stroke Outcome Classification: executive summary. Circulation 1998; 97(24): 2474-8.
4. Gotoh F, Terayama Y, Amano T. Development of a novel, weighted, quantifiable stroke scale: Japan stroke scale. Stroke 2001; 32(8): 1800-7.
5. Buck D, Jacoby A, Massey A, Ford G. Evaluation of measures used to assess quality of life after stroke. Stroke 2000; 31(8): 2004-10.
6. Sturm JW, Osborne RH, Dewey HM, Donnan GA, Macdonell RA, Thrift AG. Brief comprehensive quality of life assessment after stroke: the assessment of quality of life instrument in the north East melbourne stroke incidence study (NEMESIS). Stroke 2002; 33(12): 2888-94.
7. Iguchi Y, Kimura K, Watanabe M, Shibazaki K, Aoki J. Utility of the Kurashiki Prehospital Stroke Scale for hyperacute stroke. Cerebrovasc Dis 2011; 31(1): 51-6.
8. Kong TK, Lum CM, Mo KK. Development of a hierarchical activities of daily living scale for Chinese stroke patients in geriatric day hospitals. Aging (Milano) 1995; 7(3): 173-8.
9. Lai SM, Duncan PW, Keighley J. Prediction of functional outcome after stroke: comparison of the Orpington Prognostic Scale and the NIH Stroke Scale. Stroke 1998; 29(9): 1838-42.
10. Ghandehari K, Izadi Z. The Khorasan Stroke Registry: results of a five-year hospital-based study. Cerebrovasc Dis 2007; 23(2-3): 132-9.
11. Kwakkel G, Veerbeek JM, Harmeling-van der Wel BC, Van WE, Kollen BJ. Diagnostic accuracy of the Barthel Index for measuring activities of daily living outcome after ischemic hemispheric stroke: does early poststroke timing of assessment matter? Stroke 2011; 42(2): 342-6.

12. Quinn TJ, Dawson J, Walters MR, Lees KR. Exploring the reliability of the modified rankin scale. Stroke 2009; 40(3): 762-6.

13. Saver JL, Filip B, Hamilton S, Yanes A, Craig S, Cho M, et al. Improving the reliability of stroke disability grading in clinical trials and clinical practice: the Rankin Focused Assessment (RFA). Stroke 2010; 41(5): 992-5.

14. Uyttenboogaart M, Stewart RE, Vroomen PC, De KJ, Luijckx GJ. Optimizing cutoff scores for the Barthel index and the modified Rankin scale for defining outcome in acute stroke trials. Stroke 2005; 36(9): 1984-7.

15. Baker K, Cano SJ, Playford ED. Outcome measurement in stroke: a scale selection strategy. Stroke 2011; 42(6): 1787-94.

16. Quinn TJ, Dawson J, Walters MR, Lees KR. Variability in modified Rankin scoring across a large cohort of international observers. Stroke 2008; 39(11): 2975-9.

17. Banks JL, Marotta CA. Outcomes validity and reliability of the modified Rankin scale: implications for stroke clinical trials: a literature review and synthesis. Stroke 2007; 38(3): 1091-6.

18. Cincura C, Pontes-Neto OM, Neville IS, Mendes HF, Menezes DF, Mariano DC, et al. Validation of the National Institutes of Health Stroke Scale, modified Rankin Scale and Barthel Index in Brazil: the role of cultural adaptation and structured interviewing. Cerebrovasc Dis 2009; 27(2): 119-22.

19. Quinn TJ, Dawson J, Walters MR, Lees KR. Reliability of the modified Rankin Scale. Stroke 2007; 38(11): e144.

20. Govan L, Langhorne P, Weir CJ. Categorizing stroke prognosis using different stroke scales. Stroke 2009; 40(10): 3396-9.

21. Young FB, Weir CJ, Lees KR. Comparison of the National Institutes of Health Stroke Scale with disability outcome measures in acute stroke trials. Stroke 2005; 36(10): 2187-92.

22. De HR, Horn J, Limburg M, Van Der Meulen J, Bossuyt P. A comparison of five stroke scales with measures of disability, handicap, and quality of life. Stroke 1993; 24(8): 1178-81.